Edge-centric Analytics in Networks

Xueqi Cheng
Vanderbilt University
Nashville, USA
xueqi.cheng@vanderbilt.edu

Abstract

Network analysis has evolved substantially, with notable advancements in node-centric and graph-centric tasks, yet the exploration of edge-centric analytics has been notably limited. This oversight is significant given the crucial role of edges in elucidating the complex relationships within networks, particularly in fields such as social network analysis, cybersecurity, and bioinformatics, where the dynamics of connections between entities are often pivotal. My doctoral research aims to address this gap by delving into the under-explored domain of edge-centric analytics, providing a foundational background that is crucial for advancing the field and enhancing the application of network theory in real-world scenarios. The significance of this research lies in its potential to open new avenues for inquiry and application across diverse disciplines where understanding the nuances of relational dynamics is essential.

CCS Concepts

• Mathematics of computing → Graph algorithms; **• Theory of computation → Social networks**.

Keywords

Edge-centric analytics, network analysis, deep learning on graphs

ACM Reference Format:
Xueqi Cheng. 2025. Edge-centric Analytics in Networks. In *Proceedings of the Eighteenth ACM International Conference on Web Search and Data Mining (WSDM '25), March 10–14, 2025, Hannover, Germany.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3701551.3707418

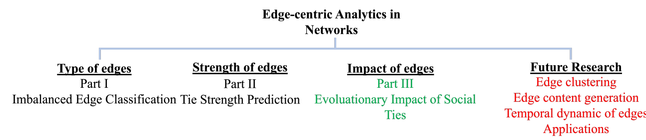

Figure 1: The paradigm of research, where back denotes previous research, green denotes current ongoing research, and red denotes possible future research.

1 INTRODUCTION AND MOTIVATION

In recent years, the field of network analysis has experienced significant progress, particularly in node-centric [27] and graph-centric tasks [28]. Despite these advancements, the domain of edge-centric analytics, crucial to understanding the relationships within networks, has remained relatively unexplored outside the confines of link prediction. This gap is particularly noticeable given the importance of edge-centric tasks in diverse areas such as social network analysis [18], cybersecurity [12], and bioinformatics [23], where the dynamics of connections between entities are often pivotal [15].

To address this gap, my doctoral research proposes to thoroughly and systematically explore edge-centric analytics within networks. The first part of my work has concentrated on edge classification, specifically targeting the challenge of topological imbalance. Subsequently, I expanded the study to an under-explored and important setting of edge classification. Specifically, we conducted a comprehensive analysis of tie strength prediction, including definitions, prediction methods, and identifying directions for future research. Then we seek to understand how various local interactions impact the macro communities by comparing the impact of dissolved social ties on the evolution of cohesive subgroups and the impact of newly formed ties. In the future, we intend to expand my research to include a broader array of edge-centric analytics tasks, including edge clustering, the analysis of edge dynamics over time, the generation of edge content, and edge-centric simulation in networks, among others. Moreover, I am keen on applying the insights gained to practical applications, such as improving relation prediction in knowledge graphs and refining the recommendation algorithms for social norms.

Overall, this work contributes a comprehensive examination of edge-centric analytics in networks, highlighting their significance across various real-world applications. Through this work, we aim to illuminate the complexities surrounding edge-centric tasks and stimulate further research and innovation in this essential yet under-explored facet of network analysis. In the remainder of this research statement, I will outline related work, delve into the details of my current and future projects, and discuss specific research questions that will be addressed during the doctoral consortium.

2 BACKGROUND AND RELATED WORK

Networks are graphical representations illustrating relationships (edges) between variables (nodes) [2]. Research in this field enables the analysis of complex patterns of relationships and structural features within networks [10]. This analysis is crucial for developing a theoretical understanding of network formation and evolution, providing valuable insights for applications in diverse domains such as social media [17], healthcare [24], and policy science [19].

Network research involves the development of tools and theory to measure and understand network properties and the application of various techniques, including heuristic-based methods and neural networks, to perform network-related tasks [4]. Over the past decades, numerous metrics have been proposed. For instance, centrality [3] metrics assess the relative importance of nodes and

WSDM '25, March 10–14, 2025, Hannover, Germany

ACM ISBN 979-8-4007-1329-3/25/03
https://doi.org/10.1145/3701551.3707418

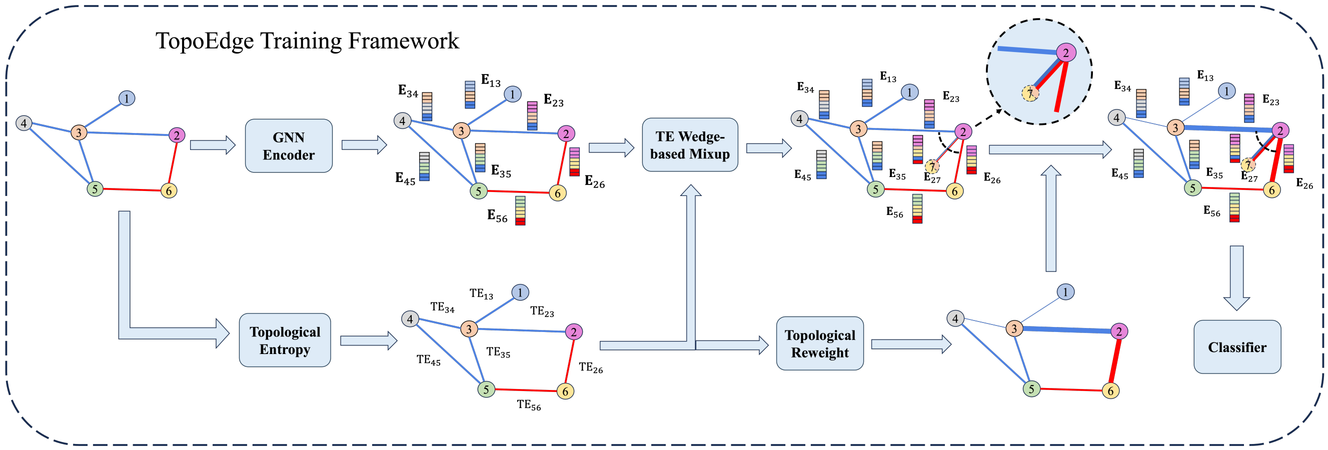

Figure 2: The overall framework of TopoEdge. The training begins with utilizing the original graph to compute edge embedding E and edge TE values. These components facilitate the execution of a TE wedge-based mixup strategy, leading to the creation of synthetic edges and their respective embeddings (i.e., E_{27} in the figure). Finally, TE values inform a topological reweighting process, producing edge weights. The ensemble of edge weights and embeddings forms the input for the final classifier.

are used for node classification tasks [16]. Modularity quantifies the division of a network into communities and aids in graph clustering [14]. Common neighborhood concepts underpin link prediction algorithms by estimating the likelihood of new connections between nodes [25]. However, most research has been focused on node/graph-centric tasks and link prediction, and research on the edge-centric analytics of networks is limited. While some initial work has addressed tasks like edge classification [1], tie strength prediction [8], and tie evolution [20], a thorough understanding of edge properties remains lacking. Hence, our research aims to conduct systematic edge-centric analytics to deepen our understanding of these elements within networks and bridge this research gap.

3 PROBLEMS AND METHODOLOGIES

Edge-centric analytics within networks encompass a range of tasks related to edge classification, tie strength prediction, evaluating the evolutionary impact of ties, and exploring their practical applications. The overall research paradigm is shown in Figure 1.

3.1 Imbalanced Edge Classification

Recent years have witnessed the remarkable success of applying Graph Machine Learning (GML) to network tasks such as node/graph classification and link prediction. However, edge classification task that enjoys numerous real-world applications, ranging from social network analysis to cybersecurity, has not seen significant advancement with the progress of this field. To address this gap, our study pioneers a comprehensive approach to edge classification. We identify a novel 'Topological Imbalance Issue,' which arises from the skewed distribution of edges across different classes, affecting the local subgraph of each edge and harming the performance of edge classifications. Inspired by the recent studies in node classification that the performance discrepancy exists with varying local structural patterns, we introduce Topological Entropy (TE), a novel topological-based metric that measures the topological imbalance for each edge. Our empirical studies confirm that TE effectively measures local class distribution variance, and indicate that prioritizing edges with high TE values can help address the issue of topological imbalance. Motivated by this observation, we develop two strategies - Topological Reweighting and TE Wedge-based Mixup - to adaptively focus training on (synthetic) edges based on their TEs. While topological reweighting directly manipulates training edge weights according to TE, our wedge-based mixup interpolates synthetic edges between high TE wedges. To further enhance performance, we integrate these strategies into a novel topological imbalance technique for edge classification: **TopoEdge** [5] (shown in Figure 2). Extensive experiments on real-world datasets demonstrate the efficacy of our proposed strategies. Additionally, our curated datasets and designed experimental settings establish a new benchmark for future edge classification research, particularly in addressing imbalance issues.

3.2 Tie Strength Prediction

During the first study, I became quite interested in the social tie strength problem, since not only has it been under-explored in recent years compared to other network analysis tasks, but it can be seen as a specific setting of edge classification. The primary difference in tie strength is that generally there lacks ground truth values due to various reasons, coming from both the lack of collection in online social networks, or even difficulty in individuals providing unbiased opinions to rank and/or classify their connections. Driven by these challenges, we explicitly conduct a comprehensive analysis of tie strength prediction w.r.t. definitions, prediction methods, and future research directions [6]. We first categorize mainstream understandings of tie strength into seven standardized definitions and verify their effectiveness by investigating the class distributions and correlations across these definitions. We also draw key insights into tie resilience from the perspective of tie dissolution that (1) stronger ties are more resilient than weaker ones, and (2) this tie resiliency ratio increases as the network evolves. We then conduct extensive experiments to evaluate existing tie strength prediction methods under these definitions, revealing that (1) neural network methods capable of learning from semantic features hold great potential for high performance, (2) models struggle under definitions that offer limited understandings of tie strength in the network, (3) existing models face imbalance issues that cannot be addressed by traditional quantity imbalance techniques, and (4) different definitions of tie strength allow for the inference of not only the current state but also the future state of a tie. Based on these findings, we propose strategies to improve existing methods and suggest promising directions for future research.

3.3 Evolutionary Impact of Social Ties

After building a stronger foundation in edge-centric analysis, we seek to understand how various local interactions impact the macro community. This leads to our current ongoing project of online polarization. Specifically, we seek to derive a comparison analysis between the impact of dissolved and newly formed ties on the evolution of cohesive subgroups [7, 26]. As a metric to evaluate the significance of newly formed/ dissolved ties on affecting the cohesive subgroups, we propose to leverage the average level of toxicity [11, 13] in the comments made to users within or between subgroups (i.e., intra- ψ_{intra} or inter- ψ_{inter} toxicity), where a larger toxicity ratio $\Psi = \psi_{\text{inter}}/\psi_{\text{intra}}$ can be one form of observing polarization between the subgroups. Then based on the evaluation metric, we would conduct several tests and comparisons to examine the impacts of dissolved/newly formed ties on increasing the toxicity in the network and thus have a deeper understanding of the formation and evolution of cohesive subgroups over time.

4 FUTURE DIRECTIONS

To broaden the scope of edge-centric analytics in networks, our future research plans include exploring a range of relatively unexamined tasks. These tasks encompass advances in research of edge clustering [22], analyzing how edges evolve, generating edge content based on characteristics, and simulating edge-centric scenarios within networks. Additionally, the insights derived from this research will be leveraged in practical applications. Specifically, we aim to enhance relational prediction within knowledge graphs and refine recommendation algorithms that are crucial for upholding social norms [9].

5 SPECIFIC RESEARCH ISSUES FOR DISCUSSION

This research aims to enhance our theoretical understanding of edges within networks and their relevance to practical edge-related tasks. A primary concern is the scarcity of high-quality real-world datasets, particularly temporal datasets that accurately reflect public opinions on significant issues. Additionally, considering the increasing influence of generative models [21], such as large language models and diffusion models, on society, it is crucial to explore how edge-centric network research can be integrated with these technologies. This discussion will 1) examine potential research directions that could combine the strengths of network analysis with the capabilities of generative models to address complex real-world problems, and 2) identify key network-related issues arising from the prevalence of generative models.

References

[1] Charu Aggarwal, Gewen He, and Peixiang Zhao. 2016. Edge classification in networks. In *2016 IEEE 32nd ICDE*. IEEE.
[2] Anastasia Bezerianos, Fanny Chevalier, Pierre Dragicevic, Niklas Elmqvist, and Jean-Daniel Fekete. 2010. Graphdice: A system for exploring multivariate social networks. In *Computer graphics forum*, Vol. 29. Wiley Online Library, 863–872.
[3] Stephen P Borgatti. 2005. Centrality and network flow. *Social networks* 27, 1 (2005), 55–71.
[4] Stephen P Borgatti, Daniel J Brass, and Daniel S Halgin. 2014. Social network research: Confusions, criticisms, and controversies. In *Contemporary perspectives on organizational social networks*. Emerald Group Publishing Limited, 1–29.
[5] Xueqi Cheng, Yu Wang, Yunchao Liu, Yuying Zhao, Charu C Aggarwal, and Tyler Derr. 2025. Edge Classification on Graphs: New Directions in Topological Imbalance. *The 18th ACM International Conference on Web Search and Data Mining (WSDM)* (2025).
[6] Xueqi Cheng, Catherine Yang, Yuying Zhao, Yu Wang, Hamid Karimi, and Tyler Derr. 2024. A Comprehensive Analysis of Social Tie Strength: Definitions, Prediction Methods, and Future Directions. *arXiv preprint arXiv:2410.19214* (2024).
[7] Kenneth A Frank. 1996. Mapping interactions within and between cohesive subgroups. *Social networks* 18, 2 (1996), 93–119.
[8] Eric Gilbert and Karrie Karahalios. 2009. Predicting tie strength with social media. In *Proceedings of the SIGCHI conference on human factors in computing systems*.
[9] Michael Hechter and Karl-Dieter Opp. 2001. Social norms. (2001).
[10] David Hevey. 2018. Network analysis: a brief overview and tutorial. *Health psychology and behavioral medicine* 6, 1 (2018), 301–328.
[11] Hossein Hosseini, Sreeram Kannan, Baosen Zhang, and Radha Poovendran. 2017. Deceiving google's perspective api built for detecting toxic comments. *arXiv preprint arXiv:1702.08138* (2017).
[12] Jiaxin Jiang, Yuan Li, Bingsheng He, Bryan Hooi, Jia Chen, and Johan Kok Zhi Kang. 2022. Spade: A real-time fraud detection framework on evolving graphs. *Proceedings of the VLDB Endowment* 16, 3 (2022), 461–469.
[13] Alyssa Lees, Vinh Q Tran, Yi Tay, Jeffrey Sorensen, Jai Gupta, Donald Metzler, and Lucy Vasserman. 2022. A new generation of perspective api: Efficient multilingual character-level transformers. In *Proceedings of the 28th ACM SIGKDD conference on knowledge discovery and data mining*. 3197–3207.
[14] Yunfei Liu, Jintang Li, Yuehe Chen, Ruofan Wu, Ericbk Wang, Jing Zhou, Sheng Tian, Shuheng Shen, Xing Fu, Changhua Meng, et al. 2024. Revisiting Modularity Maximization for Graph Clustering: A Contrastive Learning Perspective. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 1968–1979.
[15] Brian J Loasby. 2001. Time, knowledge and evolutionary dynamics: why connections matter. *Journal of Evolutionary Economics* 11 (2001), 393–412.
[16] Asmaa M Mahmoud, Abeer S Desuky, Heba F Eid, and Hoda A Ali. 2023. Node classification with graph neural network based centrality measures and feature selection. *International Journal of Electrical and Computer Engineering* (2023).
[17] Alexandra Marin and Barry Wellman. 2011. Social network analysis: An introduction. *The SAGE handbook of social network analysis* (2011), 11–25.
[18] Babita Pandey, Praveen Kumar Bhanodia, Aditya Khamparia, and Devendra Kumar Pandey. 2019. A comprehensive survey of edge prediction in social networks: Techniques, parameters and challenges. *Expert Systems with Applications* 124 (2019), 164–181.
[19] Roderick AW Rhodes. 2008. Policy network analysis. (2008).
[20] JC Rowland and WE Dietrich. 2005. The evolution of a tie channel. *River, Coastal and Estuarine Morphodynamics: RCEM* 1 (2005), 725–736.
[21] Kris Sankaran and Susan P Holmes. 2023. Generative models: An interdisciplinary perspective. *Annual Review of Statistics and Its Application* 10, 1 (2023), 325–352.
[22] Daniel K Sewell. 2020. Model-based edge clustering. *Journal of Computational and Graphical Statistics* 30, 2 (2020), 390–405.
[23] Yin-Te Tsai and Zih Yuan Lin. 2020. A survey on edge computing in bioinformatics and health informatics. In *2020 IEEE International Conference on Bioinformatics and Biomedicine (BIBM)*. IEEE, 2203–2208.
[24] Fei Wang, Uma Srinivasan, Shahadat Uddin, and Sanjay Chawla. 2014. Application of network analysis on healthcare. In *2014 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM 2014)*. IEEE, 596–603.
[25] Xiyuan Wang, Haotong Yang, and Muhan Zhang. 2023. Neural common neighbor with completion for link prediction. *arXiv preprint arXiv:2302.00890* (2023).
[26] Stanley Wasserman. 1994. Social network analysis: Methods and applications. *The Press Syndicate of the University of Cambridge* (1994).
[27] Shunxin Xiao, Shiping Wang, Yuanfei Dai, and Wenzhong Guo. 2022. Graph neural networks in node classification: survey and evaluation. *Machine Vision and Applications* 33, 1 (2022), 4.
[28] Yu Xie, Yanfeng Liang, Maoguo Gong, A Kai Qin, Yew-Soon Ong, and Tiantian He. 2022. Semisupervised graph neural networks for graph classification. *IEEE Transactions on Cybernetics* 53, 10 (2022), 6222–6235.

accepted 30 November 2024